

Jean-Christophe Kummer

Socio directivo de Noa Audio Solutions

Austria. También es un placer para mí estar aquí; no sólo como ponente, ya que hemos trabajado durante un año con la Fonoteca Nacional de México. Muchos de nuestros colegas han venido para trabajar con ustedes y estamos felices de estar aquí y ver los resultados del año de trabajo con la Fonoteca.

Les hablaré de la experiencia que hemos tenido con nuestros clientes y sobre cómo dirigir el proceso de preservación de archivos de audio. No importa desde donde veamos las cosas, la situación es la misma a nivel internacional: tenemos diferentes soportes y un total de 30,000 ó 120,000 horas de reproducción. En Austria, una persona tendría que trabajar 160,000 horas al año para terminar el proceso y aun así se tardaría del 2024 al 2084. Un proceso como éste no lo puede manejar solo un individuo que haya sido contratado para la instalación de un software. Para digitalizar todo su archivo necesitan tener un enfoque más inclusivo para la digitalización; entonces no importa en dónde estén, el reto es el mismo.

¿Cuál es el reto? Tienen que invitar a todos a participar y contemplar el flujo de trabajo y el control de calidad. Nosotros contamos con tres líneas de negocio y una solución líder para resolver el problema y mantener el control de calidad. La Fonoteca Nacional usa Job Database y también ofrecemos Ingest Line y en lo que se refiere a la preservación tenemos MediArc.

En la línea de ingesta, o Ingest Line, tenemos diferentes instrumentaciones; como pueden ver, en las estaciones tenemos hasta tres sistemas paralelos interconectables, reproductores de cintas o cualquier tipo de combinación de diferentes dispositivos de reproducción. Fabricamos nuestros propios convertidores, son 125, con control dinámico para el manejo de señales analógicas; tenemos también diferentes almacenadores de discos digitales, con una capacidad de hasta ocho discos al mismo tiempo y DAT, se puede hacer una digitalización paralela de todos estos discos. Para los medios analógicos podemos digitalizar en paralelo con dos o tres por estación.

Este es el flujo de trabajo. Aquí tenemos los sistemas de ingesta y de control de calidad. Esto quiere decir que existen algunas soluciones intermedias de activos y de metadatos —Exlibris, por ejemplo— como las bases de datos internas de donde exportamos información como archivo XML y distintos formatos hacia la solución que tenemos aquí abajo, los metadatos que se recopilan, se digitalizan en un proceso paralelo; sin embargo, ningún archivo que entra al sistema de almacenamiento lo hace sin haber sido revisado por el supervisor y para ello se proporciona una serie completa de diferentes espectrogramas analógicos en donde tenemos diferentes tipos de información: ancho de banda, azimuth, clicks mutos, overs, etc. Por ejemplo, aquí tenemos bloques de alta frecuencia y, al mismo tiempo, los instrumentos de reproducción con otros 8 DATs, pero este archivo de 40 minutos se pasa al operador de control de calidad y éste puede decir si el archivo es aceptable o si pasa al almacenamiento. Aquí tenemos algunos elementos que necesitan tratarse de forma separada y esto evita que los archivos y cintas dañados formen parte del sistema de archivo masivo.

Les hablaré, en términos generales de lo que ha logrado Noa, ya que no les puedo dar todos los detalles. Autocat es uno de nuestros productos. Nos hemos dado cuenta de que cuando tenemos que manejar contenidos en diferentes cintas en el proceso de digitalización, por ejemplo, una sinfonía grabada en una cinta y la cinta no alcanzó, se tuvo que utilizar otro dispositivo, por lo que se tienen dos segmentos y cintas diferentes.

Autocat toma estos dos segmentos, los junta y busca una correlación automática para hacer el cruce automático y empalmar las dos fuentes; como resultado crea entonces un original de muy buena calidad. Tal vez decidan aceptar o rechazar el producto. Este es sólo una parte de la solución para que vean cómo podemos enfrentar los diferentes problemas que tenemos en el mundo de la digitalización de colecciones muy grandes en los archivos.

Tenemos al Mediabutler que usa diferentes formatos de exportación y reexportación con el mismo flujo de trabajo como Exlibiris. Tenemos la base de datos interna y un subsistema que recibe toda la información de la base de datos y que envía al sistema de almacenamiento los metadatos; ustedes deciden.

Ahora algunos resultados muy comunes, generalmente éstos, como ustedes saben, los archivos tienen tres veces más necesidad de crear metadatos correctos y casi no se compara con la necesidad de digitalizar. En el caso de los procesos, nos hemos dado cuenta de que tenemos los metadatos en algún punto separado y se necesita tiempo para compilarlos. Después de este proceso podemos iniciar la digitalización y nos interesa crear un archivo que incluya los metadatos anotados y no podemos hacer nada hasta que se tengan los metadatos. Desde el punto de vista de la eficiencia, este enfoque de dos sistemas no es muy efectivo, muchos de nuestros clientes nos han dicho ¿cómo podemos cambiar las cosas para incrementar la efectividad?

Tenemos Ingest Line, que es la solución que tenemos en la Fonoteca, y también la hemos incluido como solución de audio y del flujo de trabajo, se pueden incluir muchos pasos del flujo de trabajo en esta solución. En el caso de Media Arc, la solución está simplificada, podemos trabajar con una unidad simple, pero podemos comenzar a pensar en la capacidad de la escalabilidad y podemos utilizarla en términos de la digitalización, la preservación o incrementar la escalabilidad en la velocidad por que hay más DATs para incluir a más personas. Es un sistema más abierto y no es un sistema en donde se hace una combinación de dos sistemas diferentes, de hecho esto nos permite tomar en consideración todos los puntos del flujo del trabajo para tener una participación de los diferentes departamentos, incluso al mismo tiempo.

¿Qué es Media Arc? Es una herramienta en la que tenemos que manejar diferentes soportes, nuevos y anteriores, y diferentes contenidos de producción, bases de datos y metadatos. Este bloque representa el Media Arc, en donde tenemos la base de datos Oracle para el almacenamiento. Posteriormente veremos una presentación del ejecutable, las máquinas, motores de transcodificación, el elemento de control de datos y también las interfases abiertas API. Hay clientes que desarrollan su propio sistema de transcodificación, con base en sus sistemas API, lo que nos da como resultado diferentes contenidos de audio, video y metadatos que pueden entregarse al sistema de producción; sin embargo, el contenido está completo

únicamente con un repositorio local para la ingesta. Vale la pena mencionar que Ingest Line, está diseñado para el control de calidad y este es su tamaño.

Existe la posibilidad de entrar a la mediateca y hacer una búsqueda de la base de datos, utilizar diferentes elementos de búsqueda, diferentes vínculos, darle doble clic a los resultados de la búsqueda; es como una red semántica, en donde tenemos diferentes árboles hasta llegar a los soportes digitalizados y no digitalizados. Podemos entrar a estos contenidos simplemente al pulsar estos vínculos. Estos metadatos presentan una versión completa de los mismos, cada uno de nuestros clientes puede crear su propia base de datos de metadatos. Este es un ejemplo clásico del flujo de trabajo del sistema.

La mayoría de los clientes diseñan su propio flujo de trabajo y en ocasiones necesitan de capacitación.

Ahora, los métodos de acceso. Es necesario aislar los datos del alcance de los usuarios, por lo que tenemos un ambiente para el usuario y otro para el archivo. Los usuarios solamente reciben la información cuando se tiene la copia en el archivo, nunca existe la posibilidad de que el usuario entre directamente al archivo. Hay que aislar el archivo completamente a través de una herramienta: Storage Connector, que utiliza la unión de diferentes sistemas. Comenzamos un sistema así, aquí, con base en la tecnología Noa.

Hablemos acerca de una de estas experiencias, el caso de Eslovenia; 2.5 millones de personas habitan ese país y 2,000 de estas personas trabajan en la industria de la radio y la televisión. La industria y la proporción son increíbles, tienen dos estaciones locales, una en Koper y otra en Aouba, lo que necesitan es un repositorio central en Ljubljana y de esta forma tener acceso completo a los archivos a través de las estaciones regionales.

Vamos a ver cómo funciona este concepto y espero no les impacte esta fotografía.

Hemos hablado acerca del entorno del usuario, aquí tenemos al usuario del entorno de la versión simplificada y la versión más detallada. La estación en Koper, que tiene una interfase Web, usa el servidor para tener acceso a la base de datos. Si un usuario en Koper decide que quiere archivos, y dichos archivos ya se han replicado en su estación en la caché local, pudiesen decir: “bueno, yo necesito un segmento de 6 minutos, pero yo quiero este segmento que ya está en MPEG Nivel 2”. Este extracto sale de una herramienta que haya hecho la copia del archivo lineal, el sistema efectúa una búsqueda del almacenamiento de datos, copia el archivo lineal, lo transcodifica a través de MediaButlers y proporciona el archivo, finalmente a este sistema, al ambiente del usuario o a la producción de Radio de Luvliana, este es un enfoque en donde el usuario descentralizado puede recibir el beneficio de los archivos creados linealmente, puede decidir, qué archivos necesita, porque puede tener acceso al archivo.

Ahora ¿qué es lo que ha sucedido? Con Media GUI es posible centralizar los metadatos, podemos introducirlos para hacer la búsqueda en lo que ya está digitalizado. Esto nos lleva al segundo concepto, quiero mencionar que ha sido muy importante en este país, en la siguiente mesa vamos a tener a Catharina, quien va a hablar acerca de su proyecto de televisión, el día de ayer nos reunimos en la hora de la comida y me di cuenta que Catharina es una de las

principales beneficiarias del sistema de este trabajo –nos reunimos aquí en México, imagínense, después de tres años–. En Estocolmo tienen cuatro entidades legales, SRF, con Media Arc y un archivo de Petabyte, un canal educativo y la televisión SVT. Ellos están en Anga, a 400 millas ahí tienen una “fabrica de digitalización” –como le llaman ellos.

El principal objetivo fue digitalizar 100,000 horas de video en tres años, otro de los beneficiarios fue el Archivo Nacional, y este es el plano que tenemos para esta solución, un poco más sofisticada. Utilizamos los datos SRF de la base de datos, este bloque está a 500 kilómetros de Estocolmo. Aquí es donde comenzamos con el flujo de trabajo con el Media GUI, se pasa a una unidad de digitalización externa con un motor que cuenta con su API. El comando le dice hay que digitalizar estos elementos; lo que nos lleva a la creación de 140 gigabites y se hace la transferencia al elemento de almacenamiento de petabytes que está aquí, tenemos que manejar seis terabytes por día, lo cual representa un sistema muy grande, finalmente qué es lo que va a suceder, estos archivos se deben entregar al dominio del usuario TV SVT en formato televisivo; entonces, dentro del sistema tenemos este modelo, un modelo muy flexible y abierto para poder manejar diferentes aspectos del proyecto.

Esta es la parte técnica, la mencionaré rápidamente, aquí arriba tenemos los convertidores de 24 canales, la codificadora, tenemos los API en donde se realiza la conversión, la transferencia de los seis terabytes al sistema de almacenamiento masivo digital en Estocolmo y pueden ver la coexistencia completa entre los flujos de radio de Media Arc, en paralelo y en tres años tuvieron que digitalizar 60,000 discos y cintas. Todo esto se debía proporcionar de forma paralela a los beneficiarios de la radio sueca, esto es para darles una idea de lo que se puede hacer con Media Arc, es importante mencionar que el propósito fue tener la digitalización en tres años, después de tres años esto sucedió, terminamos a tiempo el proceso de digitalización y estamos muy orgullosos de ello.

Este es un proyecto muy joven en Bucarest, se abrió hace dos semanas aproximadamente, se le llama Golden Fonoteca, con un proyecto de digitalización a diez años de 250,000 soportes. Hemos desarrollado con ellos un concepto para incluir las cinco estaciones regionales con base en el mismo principio.